Stat 602 Homework 2

Kiegan Rice and Nate Garton

Due 3/4/2019

Problem 1

Problem 4.3

Problem 4.4

Problem 5.1

The following are decompositions using the matrix

$$X = \begin{bmatrix} 2 & 4 & 7 & 2 \\ 4 & 3 & 5 & 5 \\ 3 & 4 & 6 & 1 \\ 5 & 2 & 4 & 2 \\ 1 & 3 & 4 & 4 \end{bmatrix}$$

Part a

The QR decomposition of X is:

$$Q = \begin{bmatrix} -0.27 & 0.57 & 0.77 & 0.04 \\ -0.54 & -0.07 & -0.13 & 0.56 \\ -0.40 & 0.37 & -0.35 & -0.71 \\ -0.67 & -0.50 & 0.10 & -0.13 \\ -0.13 & 0.53 & -0.50 & 0.40 \end{bmatrix}, \qquad R = \begin{bmatrix} -7.42 & -6.07 & -10.25 & -5.53 \\ 0 & 4.14 & 5.99 & 2.28 \\ 0 & 0 & 1.06 & -1.23 \\ 0 & 0 & 0 & 3.57 \end{bmatrix}$$

Thus, the basis for $C(X)$ will be the columns of the matrix Q.

The Singular Value Decomposition of X is:

$$U = \begin{bmatrix} -0.50 & 0.53 & 0.16 & -0.66 \\ -0.50 & -0.59 & 0.13 & -0.11 \\ -0.45 & 0.49 & -0.25 & 0.65 \\ -0.39 & -0.32 & -0.68 & -0.08 \\ -0.36 & -0.17 & 0.65 & 0.35 \end{bmatrix}, \quad D = \begin{bmatrix} 16.58 & 0 & 0 & 0 \\ 0 & 3.78 & 0 & 0 \\ 0 & 0 & 3.38 & 0 \\ 0 & 0 & 0 & 0.55 \end{bmatrix},$$

$$V = \begin{bmatrix} -0.40 & -0.43 & -0.80 & 0.12 \\ -0.44 & 0.30 & 0.18 & 0.83 \\ -0.71 & 0.45 & 0.04 & -0.54 \\ -0.37 & -0.72 & 0.57 & -0.06 \end{bmatrix}$$

Thus, the basis for $C(X)$ will be the columns of the matrix U.

Part b

We can find the eigen (spectral) decomposition of $X'X$ by calculating the eigenvalues, D^2, and the eigenvectors, which will be the columns of V:

$$D^2 = \begin{bmatrix} 274.94 \\ 14.32 \\ 11.44 \\ 0.30 \end{bmatrix}, \quad V = \begin{bmatrix} -0.40 & -0.43 & -0.80 & 0.12 \\ -0.44 & 0.30 & 0.18 & 0.83 \\ -0.71 & 0.45 & 0.04 & -0.54 \\ -0.37 & -0.72 & 0.57 & -0.06 \end{bmatrix}$$

We can find the eigen (spectral) decomposition of XX' by calculating the eigenvalues, D^2, and the eigenvectors, which will be the columns of U (rather than V as for $X'X$):

$$D^2 = \begin{bmatrix} 274.94 \\ 14.32 \\ 11.44 \\ 0.30 \end{bmatrix}, \quad U = \begin{bmatrix} -0.50 & 0.53 & 0.16 & -0.66 \\ -0.50 & -0.59 & 0.13 & -0.11 \\ -0.45 & 0.49 & -0.25 & 0.65 \\ -0.39 & -0.32 & -0.68 & -0.08 \\ -0.36 & -0.17 & 0.65 & 0.35 \end{bmatrix}$$

Part c

The best $rank = 1$ approximation to X will be $X^{*1} = U_1 diag(d_1) V_1'$, where U_1 is a matrix with the first column of U, and V_1 is a matrix with the first column of V:

$$X^{*1} = U_1 diag(d_1) V_1' = \begin{bmatrix} -0.50 \\ -0.50 \\ -0.45 \\ -0.39 \\ -0.36 \end{bmatrix} \times [16.58] \times \begin{bmatrix} -0.40 & -0.44 & -0.71 & -0.37 \end{bmatrix}$$

$$= \begin{bmatrix} 3.35 & 3.61 & 5.88 & 3.11 \\ 3.38 & 3.64 & 5.94 & 3.13 \\ 3.08 & 3.31 & 5.40 & 2.85 \\ 2.62 & 2.83 & 4.61 & 2.43 \\ 2.45 & 2.64 & 4.31 & 2.27 \end{bmatrix}$$

The best $rank = 2$ approximation to X will be $X^{*2} = U_2 diag(d_1, d_2) V_2'$, where U_2 is a matrix with the first two columns of U, and V_2 is a matrix with the first two columns of V:

$$X^{*2} = U_2 diag(d_1, d_2) V_2' = \begin{bmatrix} -0.50 & 0.53 \\ -0.50 & -0.59 \\ -0.45 & 0.49 \\ -0.39 & -0.32 \\ -0.36 & -0.17 \end{bmatrix} \times \begin{bmatrix} 16.58 & 0 \\ 0 & 3.78 \end{bmatrix} \times \begin{bmatrix} -0.40 & -0.44 & -0.71 & -0.37 \\ -0.43 & 0.30 & 0.45 & -0.72 \end{bmatrix}$$

$$= \begin{bmatrix} 2.48 & 4.20 & 6.78 & 1.66 \\ 4.34 & 2.97 & 4.95 & 4.75 \\ 2.28 & 3.86 & 6.23 & 1.51 \\ 3.16 & 2.46 & 4.07 & 3.33 \\ 2.73 & 2.44 & 4.01 & 2.74 \end{bmatrix}$$

Part d

Note that

$$\tilde{X} = \begin{bmatrix} -1 & 0.8 & 1.8 & -0.8 \\ 1 & -0.2 & -0.2 & 2.2 \\ 0 & 0.8 & 0.8 & -1.8 \\ 2 & -1.2 & -1.2 & -0.8 \\ -2 & -0.2 & -1.2 & 1.2 \end{bmatrix}$$

The Singular Value Decomposition of $\tilde{X}$ is:

$$U = \begin{bmatrix} -0.57 & 0.16 & 0.28 & -0.60 \\ -0.51 & 0.11 & 0.69 & 0.21 \\ -0.50 & -0.25 & -0.09 & 0.69 \\ 0.34 & -0.68 & -0.32 & -0.33 \\ 0.22 & 0.65 & -0.56 & 0.03 \end{bmatrix}, \quad D = \begin{bmatrix} 3.83 & 0 & 0 & 0 \\ 0 & 3.38 & 0 & 0 \\ 0 & 0 & 2.01 & 0 \\ 0 & 0 & 0 & 0.48 \end{bmatrix},$$

$$V = \begin{bmatrix} 0.34 & -0.81 & 0.45 & 0.18 \\ -0.36 & 0.18 & 0.26 & 0.87 \\ -0.57 & 0.03 & 0.68 & -0.45 \\ 0.64 & 0.56 & 0.52 & 0.00 \end{bmatrix}$$

Therefore, the principal component directions will be the columns of V, the principal components will be the inner products $z_j = \langle x_i, v_j \rangle$, the columns of:

$$z = \begin{bmatrix} -2.19 & 0.56 & -.57 & -0.29 \\ 1.95 & 0.39 & 1.40 & 0.10 \\ -1.91 & -0.84 & -0.18 & 0.33 \\ 1.30 & -2.32 & -0.65 & -0.16 \\ 0.85 & 2.21 & -1.14 & 0.01 \end{bmatrix}$$

The "loadings" of the first principal component are the first column of the matrix V,

$$V_1 = \begin{bmatrix} 0.34 \\ -0.37 \\ -0.57 \\ 0.64 \end{bmatrix}$$

Part e

Part f

Part d - standardized

Part e - standardized

Part f - standardized

Problem 2

Part a

Part b

Problem 3

Note: This is Stat 602 HW3 from 2015 (problem 21)

Problem 4

Note: This is Stat 602 HW3 from 2015 (problem 22)

Note: There is some crossover with 502X HW2 Q15(a) here.

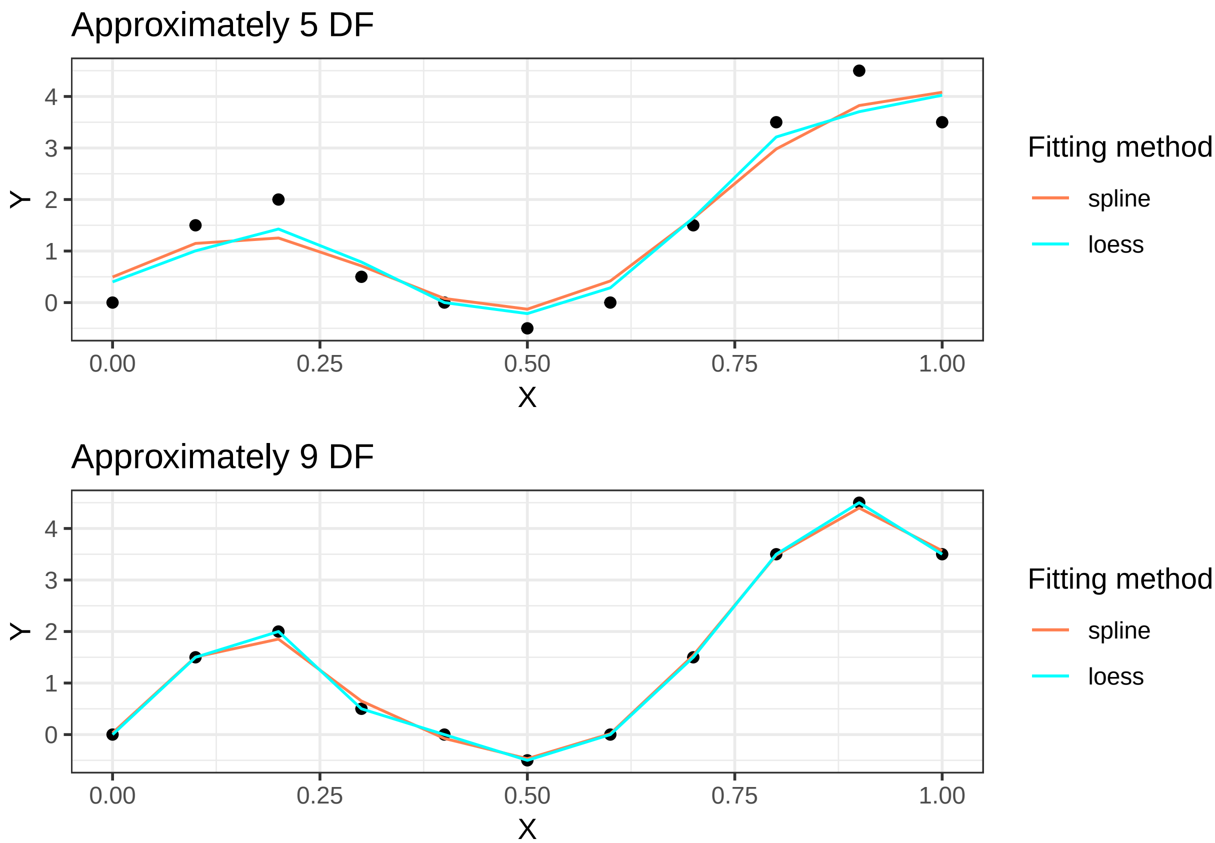

Figure 1: Plots for effective degrees of freedom 5 (top) and 9 (bottom) for Problem 6(a).

Part a

Part b

Part c

Problem 5

Note: This is Stat 602 HW3 from 2015 (problem 23)

Part a

Part b

Problem 6

Note: This is 502X HW 2 Q16(a) (**Kiegan will do this problem**)

Part a

The plots for this problem can be seen in Figure 1.

Problem 7

kNN

elastic net

PCR

PLS

MARS (in earth)

Problem 8

Part a

Part b

Part c

Part d